Synopsis for Registration as Research Student

For

Doctoral Degree

On

Deep Learning based Lung Cancer Detection and its Treatment Analysis

by

Shweta Tyagi

Research Scholar

Under the supervision of

Dr. S. N. Talbar

Department of Electronics and Telecommunication Engineering

Shri Guru Gobind Singhji Institute of Engineering and Technology,

Nanded-431606. (MS.) India.

1. Introduction

Cancer is the major cause of death worldwide. It is defined as an unregulated growth of cell which can occur anywhere in the body. Among all other types of cancer, Lung Cancer is the most common diagnosed cancer and is the leading cause of cancer deaths with incidence rate 11.6% and mortality rate 18.4%, according to GLOBOCAN report, 2018. The lung cancer cases were 2.09 million and it caused 1.76 million deaths worldwide and in India total cases were 67,795 in 2018[1]. In India, Lung Cancer is the second most common cancer among males and overall, it is ranked 4th for incidence rate and ranked 3rd for mortality rate.

Lung Cancer patients can be diagnosed with the help of different imaging tests like, X-rays, CT-scans, PET-scans etc. Lung Cancer is mostly diagnosed in its later stages when treatment becomes difficult and chances of survival are less. Due to less survival rate for lung cancer it becomes necessary to detect it at an early stage so that proper treatment plan can be designed for survival of the patient. We are working for tumor detection and its growth analysis after treatment. We have collaboration with Tata Memorial Hospital, Mumbai. We have collected data for lung tumor detection and its growth analysis. After discussion with the doctor, it is found that a treatment analysis is required to measure the tumor growth before and after treatment to find out whether the treatment is on the right path or not.

The data collected contains a total of 833 CT scans, all these can be used for tumor detection task but for treatment analysis task we have selected 612 CT scans according to baseline CT and post treatment CT for the same patient. In this task, first we will segment the tumor region for both baseline and post treatment CT, then will measure the volume and then compare the growth rate. And it can be used to examine the patient response to the treatment. This analysis will assist the doctor in deciding whether the same treatment should be continued or should switch to another plan for treatment according to the patient response to the current treatment.

1.1 Lung Anatomy

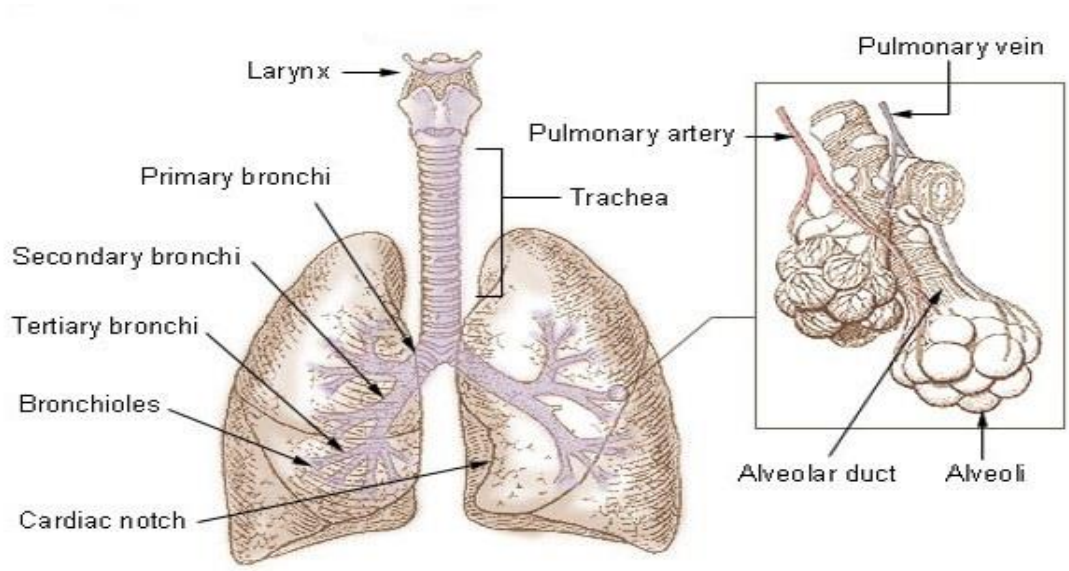

Fig.1. View of lungs and bronchi

The lungs are the primary organs for respiration and are located in the thoracic cavity separated from each other by the heart and other structures. The right lung is larger and weighs more than the left lung. Since the heart tilts to the left, the left lung is smaller than the right. And right lung has three lobes while left lung has two lobes. The lobes are separated by fissures, which are thin sheets of tissue.

Each lobe is supplied by a lobar bronchus. The lobes are subdivided by bronchopulmonary segments which are supplied by the segmental bronchi. The trachea (windpipe) conducts inhaled air into the lungs through its tubular branches, called bronchi. The bronchi then divide into smaller and smaller branches (bronchioles), finally becoming microscopic. The bronchioles eventually end in clusters of microscopic air sacs called alveoli. In the alveoli, oxygen from the air is absorbed into the blood. Carbon dioxide, a waste product of metabolism, travels from the blood to the alveoli, where it can be exhaled. The lungs are covered by a thin tissue layer called the pleura. The same kind of thin tissue lines the inside of the chest cavity -- also called pleura. A thin layer of fluid acts as a lubricant allowing the lungs to slip smoothly as they expand and contract with each breath. Each lung is invested by and enclosed in a serous pleural sac.

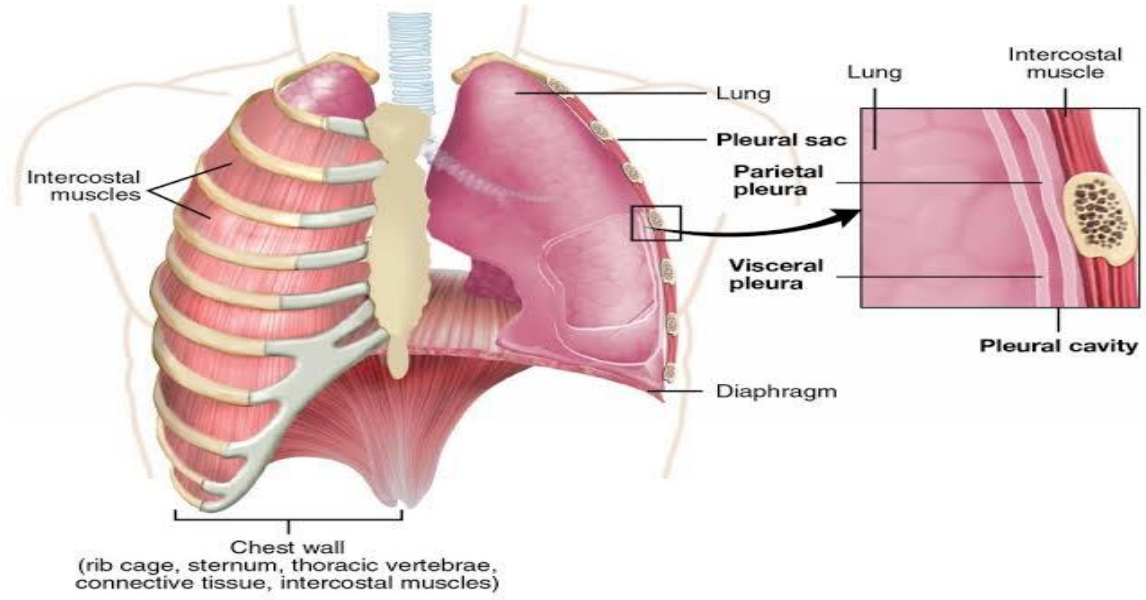

Fig.2. Pleural sac and membranes

1.2 Different Lung Conditions

- **Chronic obstructive pulmonary disease (COPD):** Damage to the lungs results in difficulty blowing air out, causing shortness of breath. Smoking is most common cause.
- **Emphysema:** A form of COPD usually caused by smoking. The fragile walls between the lungs' air sacs (alveoli) are damaged, trapping air in the lungs and making breathing difficult.
- **Pneumonia:** Infection in one or both lungs. Bacteria, especially Streptococcus pneumoniae, are the most common cause.

- **Asthma:** The lungs' airways (bronchi) become inflamed and can spasm, causing shortness of breath and wheezing. Allergies, viral infections, or air pollution often trigger asthma symptoms.
- **Pulmonary fibrosis:** A form of interstitial lung disease. The interstitium (walls between air sacs) become scarred, making the lungs stiff and causing shortness of breath.
- **Interstitial lung disease:** A collection of conditions in which the interstitium becomes diseased. Fibrosis (scarring) of the interstitium eventually results, if the process can't be stopped.

1.3 Lung Cancer types

Two main types of tumor are

- Benign tumor (non-cancerous): A benign tumor does not spread to other parts of the body from its primary site and has usually a regular shape.
- Malignant tumor (cancerous): A malignant tumor can spread to any part of the body and may have regular or uneven shape

Tumor(diameter<=30mm) → nodule

Tumor(diameter>30mm) → mass

A mass is generally considered as malignant tumor while nodule is considered as benign but a biopsy is needed for further verification.

Two main types of lung cancer are

- Small Cell Lung Cancer (SCLC)
- Non-Small Cell Lung Cancer (NSCLC)
    - Adenocarcinoma
    - Squamous cell carcinoma
    - Large cell carcinoma

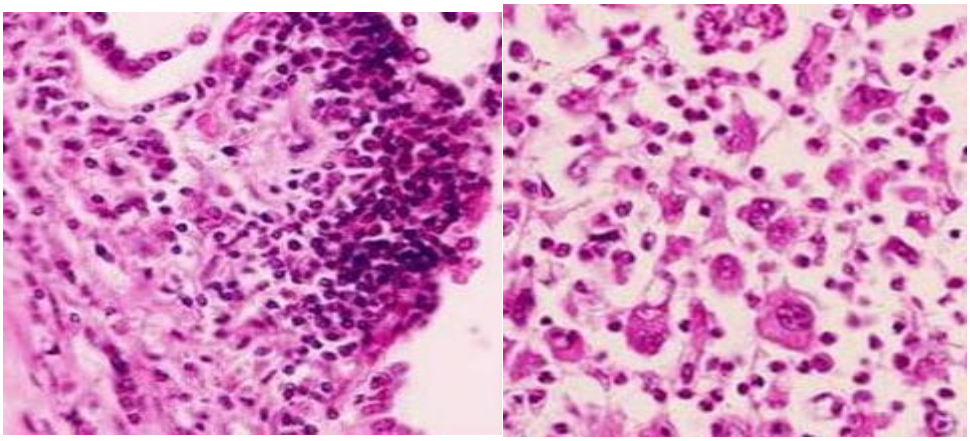

Fig.3. SCLC (left) and NSCLC (right)

This classification of is done on the basis of microscopic appearance of lung tumor cells. These two types of lung cancer are very different in nature like their growth and the extent to which they spread is quite different and they are treated in a different manner. Therefore, it is required to predict the actual type before going for treatment.

SCLC	NSCLC
Accounts for about10-15% of overall cases	Accounts for almost 85%
Microscopically, composed of much smaller cells	Comparatively larger
Often starts in bronchi	Can appear anywhere in the lungs
Grows and spreads rapidly	Grows and spreads slowly
Only 2 stages	4 stages
Mostly diagnosed in extensive stages	May be diagnosed early or in later stages
Rarely seen in non-smokers	Can be seen in non-smokers
Can be deadly in a few weeks, if not treated well	Comparatively less

Table.1. SCLC vs NSCLC

1.4 Staging of Lung Cancer

Occult stage: cancer cannot be detected through imaging tests and is also known as hidden cancer. Cancer cells can be found in mucus.

Stage 0: very small size tumor and cancer cells not spread into deeper lung tissues.

Stage I: cancer in lung tissues but not in lymph nodes.

Stage II: may have spread to lymph nodes.

Stage III: spread to lymph nodes as well as nearby organs.

Stage IV: metastasized that is spread outside the lung into other parts

Another staging of lung cancer is done on the basis of **TNM** system [3]. This staging system is most common to other cancers as well.

- **T** stands for primary **tumor**. Based on this criterion the stages are T0 (No Tumor), T1a (Tumor upto 1 cm), T1b (tumor 1-2cm), T1c (2-3cm), T2a (3-4cm), T2b (4-5cm), T3 (5-7cm) and T4 (tumor size more than 7cm).
- **N** defines whether the cancer cells have spread to the nearby lymph **nodes**. The stages are N0 (No cancer in lymph node), N1 (Cancer in lymph node within the lung), N2 (Cancer in ipsilateral lymph nodes), N3 (Cancer in contralateral lymph nodes).
- **M** denotes whether the cancer has **metastasized** (spread to other parts of the body beyond lungs) and the stages are M0 (No Metastasis), M1a (pulmonary contralateral metastasis or pleural effusion), M1b (Single metastasis in one organ outside the chest) and M1c (Multiple metastasis in one or more organs outside the chest).

1.5 Imaging Tests for Lung Cancer Detection

There are several imagining tests for diagnosing a disease like X-ray radiograph, CT-scan, PET scan, MRI, SPECT, mammogram etc. For lung cancer, initially X-rays are used. They are a form of radiation, which when pass through the body, bone and other dense objects block the radiation and look white on the film of the x-ray. The less dense tissues are hard to see and appear gray. A CT scan is a combination of a series of X-ray images taken at different angles; the CT uses a computer to create images from these X-rays. An **MRI** is a scan that uses magnetic fields and radio waves to produce a detailed image of the body's soft tissues and bones. A positron emission tomography (**PET**) scan is an imaging test which helps reveal how body tissues and organs are functioning. It uses a radioactive drug (tracer) to show this activity. This scan can sometimes detect disease before it shows up on other imaging tests. The tracer may be injected, swallowed or inhaled, depending on which organ or tissue is being studied. The tracer collects in areas of your body that have higher levels of chemical activity, which often correspond to areas of disease. On a PET scan, these areas show up as bright spots.

Mostly CT-scan is preferred for lung cancer detection because the large nodules can be easily seen on an x-ray image but nodules which are very small in size (like 1-2 mm in diameter) cannot be detected by an X-ray image. So, for that purpose CT scans are used. CT scans can give information about the specific features of the nodule, including its shape, size, location and internal density. CT scans are more accurate than chest x-rays in determining the nature of the nodule. It is less expensive and less time consuming than PET or MRI.

2. Literature Survey

2.1 Lung Segmentation and nodule detection using simple image processing

Yongbum Lee et. al [2] proposed a novel template matching technique based on genetic algorithm template matching (GATM)for detecting nodules in lung CT images. The GA was used to determine the target position in the observed image efficiently and to select an adequate template image from database. By using this method detection of nodule rate is about 72%.

Genetic Algorithms are adaptive search algorithms that belong to the larger part of evolutionary algorithms. Genetic algorithms are based on the ideas of natural selection and genetics. These are intelligent exploitation of random search provided with historical data to direct the search into the region of better performance in solution space. They are commonly used to generate high-quality solutions for optimization problems and search problems. Anita Chaudhary and Sonit Sukhraj Singh have used simple digital image processing techniques, thresholding and Marker controlled watershed segmentation for lung segmentation using MATLAB software[3]

2.2 Lung Cancer detection using some machine algorithms and classifier

M. Gomathy et al.[4] used the idea of basic image processing techniques such as morphological operations and they have used FPCM(Fuzzy Possibilistic C-Means) clustering. FPCM is used to cluster unlabelled data. R Sathishkumar et al. have used KNN algorithm and SVM classifier for lung cancer detection and classifying whether the image is normal or there is a tumor. SVM works by a separating hyper plane - a machine learning algorithm. For this algorithm, we plot data items in n dimensional space where n is the number of features with the value of the feature being equal to the value of the coordinate and then we perform classification by finding the hyper plane. SVMs are supervised learning models which analyse data for classification. They use optimum linear separating hyper planes which can be used for classification and regression. An optimum hyper plane is used to separate two sets of data in feature space and the optimum hyper plane is produced by distinguishing margins between the two sets. This means, the hyper plane will depend on border training patterns called support vectors. Here, the linear kernel SVM is used to classify the image into normal or cancerous images.

2.3 Deep Learning Models

Alex Krizhevsky et al. proposed AlexNet for Imagenet classification. Alexnet consists of 8 learned layers out of which 5 are convolutional and 3 are fully connected. The output of last fully connected layer is fed to a softmax activation function. Which provides 1000 output [7]. Karen Simonyan et al. Very Deep Convolutional Networks for Large scale Image Recognition which they called VGGNet. It consists of 16 convolutional layers and 3 fully connected layers. In hidden layers ReLU is used and at the output SoftMax is used. They have also used convolutional layer of kernel size 1 along with Convolutional filters of kernel size 3 for dimensionality reduction to reduce the number of computations [8] Olaf Ronneberger et al. proposed convolutional networks for Biomedical Image Processing. They named it as U-Net [9].

2.4 Lung Cancer Detection using Deep Learning

Wenkai Huang et al. propose noisy U-net for detection of lung cancer nodules using CT images and they have used leaky ReLU activation function [5]. Yanfeng Li et al. have used lung MRI images and used R-CNN for nodule detection [6]. Jue Jiang et al. used MRNN (multiple resolution residual networks) for lung cancer detection using CT images. Wangxia Zuo et al. used multiresolution CNN for lung nodule classification using CT images [6]. Tang et al. [13] proposed a two-stage framework for lung segmentation in chest X-rays where a segmentation model is first, they trained on 280 real images, and then fine-tuned using 5000 X-rays.

Tajbakhsh et al. [14] have used prediction of image orientation, where the input image is flipped or rotated and then the network is trained to predict this type of transformation. The authors proved that this task is very effective for lung segmentation in chest CT scans. Jin et al. [15] trained an image in-painting model based on a GAN which can generate pleural nodules in the nodule-free CT scans. The authors tested this segmentation model using 34 images having peripheral nodules from the LIDC dataset, showing that the model trained on synthetic data can achieve 2 points increase in Dice as compared to the model trained on real images

2.5 Lung Tumor Growth analysis

Mizuki Nishino et al. [16] proposed a study to measure the growth rate of tumor in ALK-rearranged advanced non-small cell lung cancer treated with crizotinib drug. They have used tumor volume segmentation software in which lung lesion was selected manually which was then segmented by the software. After segmentation, the boundaries were corrected by the observer. After manual correction of the segmentation mask boundaries, tumor volume was calculated by the software. In this study they have included 44 patients and total of 231 volume measurements were taken. In this study the tumor region was manually selected and after the segmentation done by the software again it was manually corrected by the observer. The author stated high reproducibility with interobserver concordance correlation coefficient of 0.990. Although the chances of error were less, but it is a very laborious task. We can design an algorithm which can perform faster.

2.6 Lung Dataset

Adrien Depeursinge et al[10] designed ILD lung dataset. They have collected data from 128 patients with one of the ILds and provided annotated images and a comprehensive set of 99 clinical parameters. Michael F. McNitt-Gray et al[11] designed a dataset for lung which is popularly known as LIDC dataset Seven academic centers and eight medical imaging companies collaborated to create this data set which contains 1018 cases. Three categories ("nodule > or =3 mm," "nodule <3 mm," and "non-nodule > or =3 mm") are defined by radiologists. Jo˜ao Pedrosa et al.[12] have collected lung dataset which they named as LNDb dataset. The LNDb dataset contains 294 CT scans collected retrospectively at the Centro Hospitalar e Universitário de São João (CHUSJ) in Porto, Portugal between 2016 and 2018.

3. Objectives of Research

- To detect Lung tumor in CT-scan images.
- Segmentation of lung tumor.
- Tumor growth measurement before and after the treatment to analyse the effect of treatment on a patient.
- Designing an online database.

Our main objective is to design a Deep Learning algorithm to perform all these tasks of lung tumor detection, segmentation and its treatment analysis so that we can assist the doctor in making a decision about the patient diagnosis and treatment quickly and in a better way and can contribute to public health care system.

The ability to comprehend and analyse the large datasets is very essential for medical decision making. Our Deep Learning Algorithm will provide a complementary tool for medical experts. So, the labor work will be reduced, and the overall process will speed up. This will reduce the time required by the doctor to decide the specific treatment plan for individual patients which will help to improve the survival rate.

4. Relevance to Society

Department of Science and technology (DST) has recently proposed a draft of Scientific Social Responsibility (SSR) policy. The main aim of SSR policy is to connect science with society. Our Deep Learning based algorithm can reduce the burden on radiologists and help the medical experts by proving second opinion to make a clinical decision about the patient's treatment which will help to increase the survival rate and improve the overall health of lung cancer patients. Our project is a small step towards achieving this goal of SSR policy of connecting science with society, as it will help the Doctors in making a clinical decision about the patients' treatment and to improve their survival chances which will help the lung cancer patients in achieving a healthy life.

India has improved its ranking in Global Innovation Index (GII). The theme of GII 2019 was "Creating Healthy Lives: The Future of Medical Innovation". Our work is in Medical Image Analysis and we are working towards Creating Healthy Lives for Lung Cancer patients. Therefore, we will contribute some good work and help the country o further improve its ranking in GII

5. Research Methodology

- Artificial Intelligence (AI) and Machine Learning have proved their significance by providing promising results in various fields including the field of Medical Image Processing. Deep learning is a Machine Learning process in which deep hidden layers are used to extract the high-level features from a large dataset which improve the accuracy of highly complex processes. Deep learning performs the task of data Mining very efficiently. Data Mining is a process of identifying patterns in very large datasets to predict the outcome
- For this research we will be using Deep Learning Algorithm. A Deep Learning Algorithm requires large dataset for training the model to get higher accuracy. The lung CT image datasets are publicly available on online platforms such as LIDC, LUNA 16, LNDb etc. for research purpose. These datasets vary in terms of different specifications like type of CT machine, amount of radiation dose used for acquiring CT images, the nature of image acquisition etc.
- These publicly available datasets do not include data from Indian Cancer Hospitals. Due to data variability there is a need to develop a robust algorithm which can perform efficiently on Indian Lung Cancer datasets. For this purpose, we are collaborating our research work with Tata Memorial Cancer Hospital, Mumbai. We have collected data from there to perform an analysis for overall lung cancer patients including patients from India.

- There are several deep learning models available online. For our research work, we can either fine tune any one of the models available or we can design our own model, according to whichever process provides higher accuracy and performance.

6. Work Plan

The duration of the research is 3 years so work plan accordingly is as follows

S.No.	Time period	Work plan
1.	For 1st year	• To learn about Deep Learning and Neural Networks through online courses and literature available • To study the literature for deep learning models and lung cancer detection • To perform the task of tumor detection and segmentation using pretrained models.
2.	For 2nd year	• To perform the task of tumor detection and segmentation using model trained by our own data • To learn about how to annotate the data which further we can verify with the help of doctors • To perform the treatment analysis • Writing the paper for the tasks performed
3.	For 3rd year	• To complete all the remaining tasks of the research and may be working for the further tasks related to Lung cancer

Table.2. Work plan

7. Conclusion

Our main aim is to design an algorithm or fine tune the existing deep learning algorithm for performing all the tasks of tumor detection, segmentaion and treatment analysis, which can assist the doctor in making a suitable and quick decision.

8. References

1) F.Bray,J.Ferlay,I.Soerjomataram,R.L.Siegel,L.A.Torre,andA.Jemal, ‘‘Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,’’ CA, Cancer J. Clinicians, vol. 68, no. 6, pp. 394–424, 2018

2) Yongbum Lee, Takeshi Hara, Hiroshi, shigeki and Takeo: IEEE Transactions on Medical Iamaging vol 20 No 7, July 2001.

3) Anita Chaudhary and Sonit Sukhraj Singh “Lung Cancer Detection on Ct Images By Using Image Processing”, IEEE, International Conference on Computing Sciences, (2012) pp. 142-146

4) Wenkai Huang and Lingkai Hu, “Using a Noisy U-Net for Detecting Lung Nodule Candidates”, IEEE Access vol. 7, May 2019 pp. 67905-67915

5) Yanfeng Li, Linlin Zhang, Houjin Chen, and Na Yang, “Lung Nodule Detection With Deep Learning in 3D Thoracic MR Images”, IEEE Access vol. 7, March 2019 pp. 37822-37832

6) Jue Jiang, Yu-Chi Hu, Chia-Ju Liu, Darragh Halpenny, Matthew D. Hellmann, Joseph O. Deasy, Gig Mageras, and Harini Veeraraghavan, “Multiple Resolution Residually Connected Feature Streams for Automatic Lung Tumor Segmentation From CT Images”, IEEE Transactions on Medical Imaging, vol. 38, issue: 1, January 2019 pp. 134-144

7) Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton “ImageNet classification with deep convolutional neural networks”, in CACM, 2012

8) K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition”, arXiv preprint arXiv:1409.1556,

9) Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, arXiv:1505.04597v1 [cs.CV] , 2015

10) Depeursinge, Alejandro Vargas,Alexandra Platon, Antoine Geissbuhler, Pierre-Alexandre Poletti, Henning Muller, “Building a Reference Multimedia Database for Interstitial Lung Diseases”, Elsevier, Computer Medical Imaging and Grahics vol.36 pp 227-238,2012

11) Michael F. McNitt-Gray, Samuel G. Armato III, Charles R. Meyer, Anthony P. Reeves, Geoffrey McLennan, Richie C. Pais, John Freymann, Matthew S. Brown, Roger M. Engelmann, Peyton H. Bland, Gary E. Laderach, Chris Piker, Junfeng Guo, Zaid Towfic, David P-Y. Qing, David F. Yankelevitz, Denise R. Aberle, Edwin J. R. van Beek, Heber MacMahon, Ella A. Kazerooni, Barbara Y. Croft, and Laurence P. Clarke, “The Lung Image Database Consortium (LIDC) Data Collection Process for Nodule Detection and Annotation”, Pubmed 2011

12) Joao Pedrosa, Guilherme Aresta, Carlos Ferreira, M´arcio Rodrigues, Patr´ıcia Leit˜ao, Andr´e Silva Carvalho, Jo˜ao Rebelo, Eduardo Negr˜ao, Isabel Ramos, Ant´onio Cunha, and Aur´elio Campilho, “LNDb: A Lung Nodule Database on Computed Tomography”, arXiv:1911.08434v3 [eess.IV] 19 Dec 2019

13) Tang Y., Tang, Y., Xiao, J., Summers, R.M. “Xlsor: A robust and accurate lung segmentor on chest x-rays using criss-cross attention and customized radiorealistic abnormalities generation”, arXiv preprint arXiv:1904.09229, 2019a

14) Tajbakhsh, N., Hu, Y., Cao, J., Yan, X., Xiao, Y., Lu, Y., Liang, J., Terzopoulos, D., Ding, X., “Surrogate supervision for medical image analysis: Effective deep learning from limited quantities of labeled data”, arXiv preprint arXiv:1901.08707, 2019a.

15) Jin, D., Xu, Z., Tang, Y., Harrison, A.P., Mollura, D.J., "Ct-realistic lung nodule simulation from 3d conditional generative adversarial networks for robust lung segmentation", International Conference on Medical Image Computing and Computer-Assisted Intervention, 2018, Springer. pp. 732–740

16) Mizuki Nishino, Tomoyuki Hida, Sasha Kravets, Suzanne E. Dahlberg, Christine A , Lydon, Hiroto Hatabu, Bruce E. Johnson, and Mark M. Awad, "Tumor volume dynamics and tumor growth rate in ALK-rearranged advanced non-small-cell lung cancer treated with crizotinib," Elsevier European Journal of Radiology Open 2020

17) Zongwei Zhou, Vatsal Sodha, Md Mahfuzur Rahman Siddiquee, Ruibin Feng, Nima Tajbakhsh, Michael B. Gotway and Jianming Liang, "Models Genesis: Generic Autodidactic Models for 3D Medical Image Analysis," International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2019

Shweta Tyagi
Research Scholar

Dr. S. N. Talbar
Supervisor